

AZ RDA ÉS A SZEMANTIKUS ELEMKÉSZLETEK

VASS ANNA JOHANNA

Magyar Művészeti Akadémia Művészetelméleti és Módszertani Kutatóintézet,
könyvtárvezető

TARTALMI ÖSSZEFOGLALÓ

A szemantikus webbel kapcsolatban az egyik leggyakoribb állítás, hogy nem az ember, hanem a gép számára olvasható módon fogja tárolni az adatokat.

Hogyan lesz ebből „emberközpontúság”? Milyen módon tudja szolgálni a szemantikus web technikájának eltávolodása az „ember által olvasható adatoktól” „a gépek által olvasható adatok felé” az emberi megismerést? Paradox módon az információknak a weben megvalósuló „láthatóvá válása” által.

Az állítások megfogalmazása után, a könyvtári területről részben kitekintve, a legváltozatosabb gyűjteményegységek – zeneművek, régi könyvek, digitális könyvtárak, színház- és előadóművészet, jazzfelvételek, videómegosztó alkalmazás stb. – megközelítéséből hozunk példákat arra, hogy hogyan válnak feltárhatóvá az újabb eszközökkel – például az RDA által – a jövőben a kulturális örökség tartalmaira vonatkozó információk.

Bevezetés

A szemantikus webbel kapcsolatban – bár még messze van attól, hogy általános, mindennapi gyakorlattá váljon –, annyit már olvashattunk, hallhattunk, hogy nem az ember, hanem a gép számára olvasható módon fogjuk tárolni az adatokat. Hogyan lesz ebből mégis „emberközpontúság”, milyen módon tudja szolgálni a szemantikus web technikájának látszólagos eltávolodása az „ember által olvasható adatoktól” „a gépek által olvasható adatok felé” az emberi megismerést? Paradox módon az információknak a weben megvalósuló „láthatóvá válása” által. A kulturális gyűjtemények adategyüttese jellemző módon „saját” szabályok szerint, „saját” szerkezetben, a webes keresőmotorok számára azonban nehezen megközelíthető, nehezen elérhető módon tárolódnak. Így viszont az emberi tudás legjava, annak legrelevánsabb része marad rejtve, hiszen a tudásvagyon letéteményesei – még mindig – a kulturális tartalmakat őrző közgyűjtemények.

Az állítások megfogalmazása után, a könyvtári területről részben kitekintve, a legváltozatosabb gyűjteményegységek – zeneművek, régi könyvek, digitális könyvtárak, színház- és előadóművészet, jazzfelvételek, videómegosztó alkalmazás stb. – megközelítéséből hozunk példákat arra, hogy hogyan válnak feltárhatóvá az újabb eszközökkel – például az FR-modellek és az arra épülő RDA által – a jövőben a kulturális örökség tartalmaira vonatkozó információk.

Az adatok láthatóvá válása a weben

A digitális környezetnek meghatározóvá válásával keletkezett kihívások megjelenése az IFLA-programokban a kilencvenes évek óta egyre jobban nyomon követhető. Legfőképpen a nem hagyományos dokumentumok leírásának problémái, valamint az információ közzétételében, szervezésében jelentkező új problémák voltak azok az okok, melyek végül a meglévő katalogizálási szabványok felülvizsgálatához, módosításához vezettek.¹ Ezeknek a folyamatoknak a sorába illeszkedik többek között a *Functional Requirements for Bibliographic Records* (FRBR) modell, majd az annak nyomán megszülető katalogizálási szabályzat, a *Resource Description and Access* (RDA) kidolgozása is.

Ha most visszatekintünk az RDA-val kapcsolatos korai – elsősorban a 2008–2011 között született – szakirodalom várákozásaira, akkor látható, hogy egyrészt az interneten való megjelenés igénye, másrészt a nem hagyományos dokumentumokkal kapcsolatos ígéret tűnik hangsúlyosnak, mint ami a fejlesztéseket, a váltást indokolja.²

Mára már egyre bizonyosabb, hogy az interneten való jobb megjelenést a Linked Open Data technológia, az elemkészetek regisztrálása fogja leginkább biztosítani.³ Kiindulásképpen, az egyszerűség kedvéért érdemes feltételezni, hogy a szemantikus weben az adatok nagy tömege meghatározott infrastruktúrára alapozva úgy működik, mint egy nagy kapcsolt adatbázis. A W3C szintaktikai ajánlásainak alkalmazása e tekintetben alapvető fontosságú, legalább abban a mértékben, amennyire a weben megjelenő könyvtári forrásokat alkalmassá tehetik a remélt szemantikai műveletekre.⁴ A *Resource Description Framework* (RDF) szintaktikája az adatok felépítését háromtagú állításokban, ún. tripletekben határozza meg: „A” objektumhoz tartozik egy „B” tulajdonság vagy jellemző, és ennek a tulajdonságnak „C” értéke van. Minden elem, tehát a tripletek tagjai, kifejezhetők egy-egy URI-val. RDF-ben – mivel az egy generikus szerkezet, amely mára már nagyon sok eszköztárat (DC, FOAF, SKOS, BIBFRAME, FRBRoo, RDA, stb.) használ –, lényegében bármely bibliográfiai rekordban lévő bármely adatot ki lehet így fejezni az adatok valamelyes átalakításával. Az ember által olvasható címkék, elnevezések azonban a gépek számára nem érthetők, mert többértelműek lehetnek, illetve nehezen formalizálhatók, például más-más környezetben előfordulhat ugyanaz a karaktorsor más jelentéssel, és az így adódó jelentéskülönbséget csak az emberi intelligencia képes feloldani. Miután a számítógépnek nincs emberi intelligenciája, az RDF úgy aknázza ki a web szerkezetének lehetőségeit, hogy egy speciális azonosítót használ, a Uniform Resource Identifiert, vagyis URI-t, amely betűkből, számokból és írásjelekből állhat, bizonyos írásjelek viszont tiltva vannak. Az RDF előírja, hogy az alany (subject) és az állítmány (predicate vagy property) URI-val legyen azonosítva. A tárgy (object) viszont lehet URI is és szöveges rész is, vagy dátum. Az információ fő hordozója ebben a szerkezetben az állítmány, amely az alanyt és a tárgyi részt összekapcsolja. Az állítmány vagy tulajdonság azonos szerepet tölt be, mint a metaadatsémákban az attribútum. Az egyes elemekhez nemcsak URI-k, hanem címkék, megnevezések (label) is kapcsolhatók. Ez utóbbi azért

is fontos, mert az előállt adategyüttes így válik az ember számára is érthetővé, olvashatóvá. Az RDF-állítások összefűzhetők, tehát egy elem, amely az egyik állításban alany volt, a másik állításban betöltheti a tárgy szerepét, és fordítva. Összefoglalva, az RDF tehát kiaknázza a web kommunikációs képességeit, annak infrastruktúráját alkalmazva. A géppel olvasható adatok felépítését tekintve az RDF-állítás egyetlen kapcsolat, egyetlen érték és egyetlen értékkel rendelkező dolog hármasából áll, úgyhogy nevezhetjük „atomizáltnak” az általa kifejezett információcsomagot, szemben a hagyományos authority rekord tartalmával, amely ezzel a hasonlattal élve tekinthető „molekulárisnak”. Az RDF tehát részeire bontja azt az állításhalmazt, amelyet eddig authority rekordként kezeltünk.

Az itt ismertetett technológiákban rejlő lehetőségeket viszont csak akkor tudják a könyvtárak kihasználni, ha azokat az adatelemeket, elemkészleteket, amelyeket a gyűjteményeik feltárására használnak, a fent részletezett szabályok szerint közzéteszik, más szóval „regisztrálják”. A regisztrálás folyamata alapvetően az egyes elemek URI-val való azonosítását jelenti. Mára már számos könyvtári elemkészlet áll rendelkezésre ilyen módon – az FRBR, az ISBD-k, a Library of Congress tárgyszavai stb. –, azonban az RDA elemkészletének regisztrálásával 2007-ben elkezdődött fejlesztés úttörő volt ezen a téren, mintegy megnyitotta az utat a hasonló törekvések előtt.⁵

Új megközelítések – újabb gyűjtemények elérése

Az RDA fejlesztésével kapcsolatban egy helyütt olvasható, hogy „katalogizáló forradalom” zajlik.⁶ Talán túlzó ez a megállapítás, talán nem, mindenesetre az FRBR – és az annak gyakorlatba ültetéseként kezelhető RDA – megjelenésének van egy kevésbé figyelemre méltatott hatása is. Az FRBR által az adatok és erőforrások gazdaságos szervezésére, csoportosítására bevezetett *mű-kifejezési forma–megjelenési forma–példány* szintek nemcsak az elméleti irodalomra gyakoroltak hatást, hanem termékenyítőleg hatottak egy sor olyan gyűjteményegyüttessel kapcsolatos konkrét fejlesztésre is, amelyek részben kívül esnek a hagyományos könyvtári gyűjteménykezelés és feldolgozás gyakorlatán, vagy pedig olyan speciális gyűjteményrészek – „különgyűjtemények” –, amelyek hosszú ideje csak bizonyos kompromisszumok árán tudták alkalmazni a saját területükön a rendelkezésre álló katalogizálási szabályokat, illetve a könyvtári adatbázisokban használt MARC adatsere-formátum szerkezetét. A *Cataloging and Classification Quarterly* 2005-ben és 2012-ben is megjelentetett egy-egy tematikus számot, amelyet az FRBR hatásának szenteltek. Ebben a két kötetben sorra szerepelnek olyan tanulmányok, amelyek írói ihletet, inspirációt kaptak az adott területen felmerülő katalogizálási anomáliák új elméleti alapokra helyezéséhez, mint például az internetarchívumok feltárásának speciális problémái; az előadóművészet területén keletkezett alkotások adatfelvételének elmélete és gyakorlata; a népművészeti gyűjteményekben rejtve maradó oral history típusú információk megragadása; a régi könyvek katalogizálási gyakorlata; a YouTube mint mára az egyik legnagyobb, de állandó változásban lévő digitális gyűjtemény metaadatolásának

kérdései; a zenei gyűjtemények kapcsán a szerzőség, illetve az önálló mű elhatárolásának kérdései mind a klasszikus zenében, mind a jazz előadóművészet tekintetében.⁷

Az FRBR mintha mozgásba lendítette volna a könyvtári feldolgozással foglalkozó közösségeket. A felvetett problémák szélességének érzékeltetésére néhány példa:

A YouTube-bal kapcsolatban vetődik fel, de persze minden digitális gyűjteményre igaz, hogy a hozzáférés optimalizálása – avagy hogyan választják ki a felhasználók, hogy melyik videót nézzék meg, melyik a keresésnek megfelelő, jó találat, és melyik találat azonos egy másikkal – a metaadatok strukturálásának kérdéseire ráirányítják a figyelmet. A felhasználók ezrei duplikációk százait hozzák létre. Ilyen körülmények között a művek közötti kapcsolatok megfelelő szervezése, a „mű” fogalma felértékelődik, nem beszélve olyan gyűjteményspecifikus, mégis a hozzáférést is befolyásoló kérdésekről, mint hogy melyik a legjobb minőségű változat, melyiknek van hangja, mennyi memória szükséges a megtalált videó megtekintéséhez stb.

Mind az FRBR, mind a FRAD meglehetősen tág teret enged viszont a döntésnek abban a kérdésben, hogy hol húzódik a határ a *mű* és a *kifejezési forma* között, valamint, hogy mely szempontoktól függ az elsődleges szerzői felelősség, az új művet létrehozó szerzői minőség megállapítása.⁸ Az eltérés oka a kulturális meghatározottság tekintetében jelentkezik, és ez nemcsak nemzeteket, hanem eltérő felhasználói közösségeket – és gyűjteményeket – is jelenthet. Mely példán keresztül is érthetnénk meg jobban a szerzőség elhatárolásának dilemmáit, mint például a jazz zenével kapcsolatban, amelynek köztudottan az improvizáció az egyik szervező eleme, és ahol szinte minden egyes előadás az önálló invención, improvizációs készségen alapul, és ezek az előadások kétségtelenül számos esetben önálló intellektuális tartalmat hoznak létre.

A könyvtári gyűjtemények alapvetően materiális formát öltött szellemi javakkal foglalkoznak: gyűjtik, tárolják, rendszerezik az ebbe a körbe tartozó dokumentumokat. Ritkán kell azzal szembesülnünk, hogy vannak a kulturális örökségnek olyan területei, amelyek nem öltenek megragadható materiális formát, ugyanakkor jelentőségük, hatásuk az adott kultúrában nem csekély. Két, egymástól távol eső területről is lehet példát említeni. A néprajzi gyűjteményekben minden bizonnyal a világ számos pontján felmerül a szóbeli hagyománynak – imák, mesék, mondák stb. – a feltárhatósága. Ezek a „művek” leggyakrabban valaki más által leírva vagy átírva, gyűjteményes kötetekben szerepelnek, de sem önállóan, sem tárgyi melléktételként nem ragadható meg közvetlenül motivikus gazdagságuk. Az élő színpadi előadás pedig szintén megragadhatatlan az „archívumok” számára, noha számtalan módon jelen van a fizikai testet öltött, és jó esetben gyűjteménybe került – szórólapok, színházi plakátok, díszlettervek, műsorfüzetek stb. – dokumentumok által a katalógusok adatain keresztül, de mégsem „saját jogon”, önálló entitásként.

Az FRBR és az RDA lebetősegei

Valójában az elméleti modellek által soha nem lett a mű fogalma intézményesen meghatározva. Az FRBR entitásainak első csoportjában szereplő fogalmak – különösen a *mű* és a *kifejezési forma* tekintetében – egymástól való elhatárolása azóta is ismétlődően zavar forrása. A FRAD egyenesen a katalogizáló gyakorlat szerepét emeli ki a műnek mint olyannak a meghatározásában.⁹ Ez alapján elmondható, hogy jelenleg az egyik legmegosztóbb kérdés annak elhatárolása, hogy valami ugyanannak a műnek két különböző kifejeződése, vagy pedig két különböző mű-e.

Az RDA-ra mint leírási szabványra hárult az a feladat, hogy az FRBR-modellek által felvetett elméleti kategóriák egymáshoz való viszonyát letisztázza, kodifikálja, nem utolsósorban az elemkészletek regisztrálásának folyamatában. Amíg az RDA pusztán, mint szöveg létezett, ez az állapot elodázta a strukturális kérdések felvetését és megválaszolását. Az a fejlesztés azonban, amelynek során az RDA elemkészlete végül a linked data technológiával közzétételre került, felszínre hozta a szövegben rejtve maradt aszimmetriákat, illetve az FRBR-modell hiányait is.¹⁰ Az FRBR és az RDA közötti kapcsolat megteremtése kapcsán derült ki ugyanis, hogy a két rendszer között jelentős inkonzisztencia mutatkozott. Ez akkor vált nyilvánvalóvá, amikor az FRBR elméleti kategóriáit – entitások, attribútumok, kapcsolatok megnevezéseit – az RDA rendelkezéseinek kellett megfeleltetni, sőt, műveletek végzésére alkalmas elemkészletként, szótárelemként definiálni. A kiindulópont a regisztrálás során az volt, hogy egy-egy „element” az elemkészletben egy-egy FRBR entitásnak felel meg. Ugyanakkor az RDA-ban egy elem valójában egynél több FRBR entáshoz is kapcsolható, ez kimutatható például az RDA J Appendixében. A regisztrálás folyamatában el kellett továbbá dönteni, hogy a kapcsolatok mely szinten legyenek definiálva, ugyanis ez sem az FRBR-ből, sem az RDA-ból nem derül ki egyértelműen. Ugyanakkor, ha az RDA jelenlegi elemkészletét¹¹ és az RDA Toolkiten¹² keresztül elérhető szöveges változatot összevetjük, láthatjuk, hogy a változás az elemkészletben ma is folyamatos. És itt kell azt is megjegyezni, hogy az IFLA által 2017 augusztusában megjelentetett *Library Reference Model (LRM)*, amely az addigi FR-modellek konszolidált változatának tekinthető, a későbbiekben az RDA egész elemkészletére hatással lesz, amint ezt az RDA Toolkit átalakítását célzó *3R Project* bejelentésekor előrevetítették.¹³

Az RDA fejlesztésének folyamatából itt felvillantott pillanatkép valójában arra alkalmas, hogy a változások állandóságára hívja fel a figyelmet. Az új modellek megjelenése a katalogizálási szemlélet és gyakorlat átalakítását célozva való(já)ban számos kreatív és újító gondolatot inspirált a megszokottól eltérő gyűjtemények feltárásával kapcsolatban is, az eldöntetlenül maradt kérdések megválaszolásában pedig sokkal nagyobb szerep jut a mindennapi gyakorlat során az adatfelvételben résztvevő szakembereknek. Hogy forradalom zajlik-e a katalogizálás terén, ezt talán még nem tudjuk megválaszolni, mindenesetre az intellektuális kihívás folyamatos lesz az újfajta szabványkörnyezetben és az új technológiák alkalmazása során.

Jegyzetek

1. OLIVER, Chris: Changing to RDA. = *Feliciter*, 53. vol. 7. no. 2007. 250–253. p. Forrás: <http://www.rda-jsc.org/archivedsite/docs/felicitervol53no7p250-253.pdf> [2018. július 27.]
2. Lásd az előbbi jegyzetet, valamint többek közt: EL-SHERBINI, Magda – CURRAN, Megan: Resource Description and Access „RDA”: New code for cataloging. = *Serials Librarian*, 60. vol. 1–4. no. 2011. 7–15. p. Forrás: <https://www.tandfonline.com/doi/abs/10.1080/0361526X.2011.556425> [2018. július 27.]
3. Részletesebben lásd: VASS Johanna: Az RDA kapcsolata a nemzetközi könyvtári modellekkel, és elemkészségekkel II. Az RDA szótárai – tovább az FRBRoo modell felé. = *Tudományos és Műszaki Tájékoztatás*, 64. évf. 10. sz. 2017. 477–495. p.
4. GRADMANN, Stefan: fdfs:frbr – towards an implementation model for library catalogs using semantic web technology. = *Cataloging and Classification Quarterly*, 39. vol. 3–4. no. 2005. 66–67. p. Forrás: <https://doi.org/10.1300%2Fj104v39n0305> [2018. július 27.]
5. PHIPPS, John – DUNSIRE, Godon – HILLMANN, Diane: Building a platform to manage RDA Vocabularies and data for an international linked data world. = *Journal of Library Metadata*, 15. vol. 3–4. no. 2015. 253. p. <https://doi.org/10.1080%2F19386389.2015.1099990> [2018. július 27.]
6. HUWE, Terence K.: Building digital libraries. Creating the conditions for innovation. = *Computers and Libraries*, 34. vol. 6. no. 2014. 16–23. p.
7. ALBERTSEN, Ketil – NUYS, Carol van: Paradigma: FRBR and Digital Documents. = *Cataloging and Classification Quarterly*, 39. vol. 3–4. no. 2005. 125–149 p. https://doi.org/10.1300%2Fj104v39n03_09 [2018. július 27.]; Le BOEUF, Patrick: Musical works in the FRBR Model or „Quasi la stessa cosa”. Variations on a theme by Umberto Eco. Ua. 103–124. p. https://doi.org/10.1300%2Fj104v39n03_08; JONSSON, Gunilla: Cataloguing of hand press materials and the concept of expression in FRBR, Ua. 77–86. p. https://doi.org/10.1300%2Fj104v39n03_06; MILLER, David – LE BOEUF, Patrick: „Such stuff as dreams are made on”. How does FRBR fit performing arts? Ua. 151–178. p. Forrás: https://doi.org/10.1300%2Fj104v39n03_10; NICOLAS, Yann: Folklore requirements for bibliographic records. Oral traditions and FRBR, Ua. 179–195. p. Forrás: https://doi.org/10.1300%2Fj104v39n03_11; GREENBERG, Jane – TRUJILLO, Shaun – MAYER-PATEL, Ketan: YouTube: Applying FRBR and exploring the multiple description coding compression model. = *Cataloging and Classification Quarterly*, 50. vol. 5–7. no. 2012. 742–762. p. Forrás: <https://doi.org/10.1080%2F01639374.2012.681273>; SCHMIDT, Raymond: Composing in real time. Jazz performances as „Works” in the FRBR Model, Ua. 653–669. p. Forrás: <https://doi.org/10.1080%2F01639374.2012.681601>
8. A bibliográfiai tételek funkcionális követelményei : záró-jelentés / készítette az IFLA Bibliográfiai Tételek Funkcionális Követelményei Munkacsoportja ; jóváhagyta az IFLA Katalógizáló Szekciójának Állandó Bizottsága ; fordította Berke Barnabásné. – [Budapest : Országos Széchényi Könyvtár, 2006]. Forrás: <https://www.ifla.org/files/assets/cataloguing/frbr/frbr-hu.pdf> [2018. január 30.] 24. p.
9. Functional Requirements for Authority Data (2009) by IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Forrás: <https://www.ifla.org/publications/functional-requirements-for-authority-data> [2018. január 30.] 10. p.

10. Részletesen lásd erről: HILLMANN, Diane: Registering the RDA Vocabularies [ppt]. Elhangzott: 25th Annual meeting of the Authority Control konferencián, 2009. július 11-én. Forrás: <https://www.slideshare.net/smartbroad/registering-the-rda-vocabularies-1734427> [2018. január 30.] és PHIPPS, John – DUNSIRE, Godon – HILLMANN, Diane: Building a platform to manage RDA Vocabularies, and data for an international linked data world. = Journal of Library Metadata, 15. vol. 3–4. no. 2015. 252–264. p.
11. RDA Registry. Forrás: <http://www.rdaregistry.info/> [2018. február 28.]
12. RDA Toolkit. Forrás: <http://www.rdatoolkit.org/> [2018. február 28.]
13. RDA Toolkit Restructure and Redesign Project. Forrás: <http://www.rdatoolkit.org/3Rproject> [2018. január 30.]

Vass Anna Johanna 2016 márciusa óta a Magyar Művészeti Akadémia könyvtárának vezetője. Az ELTE Irodalomtudományi Doktori Iskola Könyvtártudományi Doktori Programjának hallgatója, PhD-témája: Resource Description and Access. Érdeklődési területe: a könyvtári szabványok alkalmazása a gyakorlatban.